1

Introduction

1.1 Motivation: Stochastic Differential Equations

Stochastic Integration and Stochastic Differential Equations (SDEs) appear in analysis in various guises. An example from physics will perhaps best illuminate the need for this field and give an inkling of its particularities. Consider a physical system whose state at time t is described by a vector X_t in $\mathbb{R}^n$. In fact, for concreteness' sake imagine that the system is a space probe on the way to the moon. The pertinent quantities are its location and momentum. If x_t is its location at time t and p_t its momentum at that instant, then X_t is the 6-vector (x_t, p_t) in the phase space $\mathbb{R}^6$. In an ideal world the evolution of the state is governed by a differential equation:

$$\frac{dX_t}{dt} = \begin{pmatrix} dx_t/dt \\ dp_t/dt \end{pmatrix} = \begin{pmatrix} p_t/m \\ F(x_t, p_t) \end{pmatrix} .$$

Here m is the mass of the probe. The first line is merely the definition of p: momentum = mass × velocity. The second line is Newton's second law: the rate of change of the momentum is the force F. For simplicity of reading we rewrite this in the form

$$dX_t = a(X_t)\, dt \,, \tag{1.1.1}$$

which expresses the idea that the change of X_t during the time-interval dt is proportional to the time dt elapsed, with a proportionality constant or *coupling coefficient* a that depends on the state of the system and is provided by a model for the forces acting. In the present case $a(X)$ is the 6-vector $(p/m, F(X))$. Given the initial state X_0, there will be a unique solution to (1.1.1). The usual way to show the existence of this solution is Picard's iterative scheme: first one observes that (1.1.1) can be rewritten in the form of an *integral equation*:

$$X_t = X_0 + \int_0^t a(X_s)\, ds \,. \tag{1.1.2}$$

Then one starts Picard's scheme with $X_t^0 = X_0$ or a better guess and defines the iterates inductively by

$$X_t^{n+1} = X^0 + \int_0^t a(X_s^n)\, ds \,.$$

If the coupling coefficient a is a Lipschitz function of its argument, then the *Picard iterates* X^n will converge uniformly on every bounded time-interval and the limit X^∞ is a solution of (1.1.2), and thus of (1.1.1), and the only one. The reader who has forgotten how this works can find details on pages 274–281. Even if the solution of (1.1.1) cannot be written as an analytical expression in t, there exist extremely fast numerical methods that compute it to very high accuracy. Things look rosy.

In the less-than-ideal real world our system is subject to unknown forces, *noise*. Our rocket will travel through gullies in the gravitational field that are due to unknown inhomogeneities in the mass distribution of the earth; it will meet gusts of wind that cannot be foreseen; it might even run into a gaggle of geese that deflect it. The evolution of the system is better modeled by an equation

$$dX_t = a(X_t)\,dt + dG_t\;, \tag{1.1.3}$$

where G_t is a noise that contributes its differential dG_t to the change dX_t of X_t during the interval dt. To accommodate the idea that the noise comes from without the system one assumes that there is a *background noise* Z_t – consisting of gravitational gullies, gusts, and geese in our example – and that its effect on the state during the time-interval dt is proportional to the difference dZ_t of the *cumulative noise* Z_t during the time-interval dt, with a proportionality constant or *coupling coefficient* b that depends on the state of the system:

$$dG_t = b(X_t)\,dZ_t\;.$$

For instance, if our probe is at time t halfway to the moon, then the effect of the gaggle of geese at that instant should be considered negligible, and the effect of the gravitational gullies is small. Equation (1.1.3) turns into

$$dX_t = a(X_t)\,dt + b(X_t)\,dZ_t\;, \tag{1.1.4}$$

in integrated form $$X_t = X^0_t + \int_0^t a(X_s)\,ds + \int_0^t b(X_s)\,dZ_s\;. \tag{1.1.5}$$

What is the meaning of this equation in practical terms? Since the background noise Z_t is not known one cannot solve (1.1.5), and nothing seems to be gained. Let us not give up too easily, though. Physical intuition tells us that the rocket, though deflected by gullies, gusts, and geese, will probably not turn all the way around but will rather still head somewhere in the vicinity of the moon. In fact, for all we know the various noises might just cancel each other and permit a perfect landing.

What are the chances of this happening? They seem remote, perhaps, yet it is obviously important to find out how likely it is that our vehicle will at least hit the moon or, better, hit it reasonably closely to the intended landing site. The smaller the noise dZ_t, or at least its *effect* $b(X_t)\,dZ_t$, the better we feel the chances will be. In other words, our intuition tells us to look for a statistical inference: from some reasonable or measurable assumptions on the background noise Z or its effect $b(X)dZ$ we hope to conclude about the likelihood of a successful landing.

This is all a bit vague. We must cast the preceding contemplations in a mathematical framework in order to talk about them with precision and, if possible, to obtain quantitative answers. To this end let us introduce the set Ω of all possible *evolutions of the world*. The idea is this: at the beginning $t = 0$ of the reckoning of time we may or may not know the *state-of-the-world* ω_0, but thereafter the course that the history $\omega : t \mapsto \omega_t$ of the world actually will take has the vast collection Ω of evolutions to choose from. For any two possible courses-of-history[1] $\omega : t \mapsto \omega_t$ and $\omega' : t \mapsto \omega'_t$ the state-of-the-world might take there will generally correspond different cumulative background noises $t \mapsto Z_t(\omega)$ and $t \mapsto Z_t(\omega')$. We stipulate further that there is a function $\mathbb{P}$ that assigns to certain subsets E of Ω, the *events*, a *probability* $\mathbb{P}[E]$ that they will occur, i.e., that the actual evolution lies in E. It is known that no reasonable probability $\mathbb{P}$ can be defined on *all* subsets of Ω. We assume therefore that the collection of all events that can ever be observed or are ever pertinent form a σ-algebra $\mathcal{F}$ of subsets of Ω and that the function $\mathbb{P}$ is a probability measure on $\mathcal{F}$. It is not altogether easy to defend these assumptions. Why should the observable events form a σ-algebra? Why should $\mathbb{P}$ be σ-additive? We content ourselves with this answer: there is a well-developed theory of such triples $(\Omega, \mathcal{F}, \mathbb{P})$; it comprises a rich calculus, and we want to make use of it. Kolmogorov [57] has a better answer:

Project 1.1.1 *Make a mathematical model for the analysis of random phenomena that does not require σ-additivity at the outset but furnishes it instead.*

So, for every possible course-of-history[1] $\omega \in \Omega$ there is a *background noise* $Z_. : t \mapsto Z_t(\omega)$, and with it comes the *effective noise* $b(X_t)\,dZ_t(\omega)$ that our system is subject to during dt. Evidently the state X_t of the system depends on ω as well. The obvious thing to do here is to compute, for every $\omega \in \Omega$, the solution of equation (1.1.5), to wit,

$$X_t(\omega) = X_t^0 + \int_0^t a(X_s(\omega))\,ds + \int_0^t b(X_s(\omega))\,dZ_s(\omega)\,, \qquad (1.1.6)$$

as the limit of the Picard iterates $X_t^0 \stackrel{\text{def}}{=} X_0$,

$$X_t^{n+1}(\omega) \stackrel{\text{def}}{=} X_t^0 + \int_0^t a(X_s^n(\omega))\,ds + \int_0^t b(X_s^n(\omega))\,dZ_s(\omega)\,. \qquad (1.1.7)$$

Let T be the time when the probe hits the moon. This depends on chance, of course: $T = T(\omega)$. Recall that x_t are the three spatial components of X_t.

[1] The redundancy in these words is for emphasis. [Note how repeated references to a footnote like this one are handled. Also read the last line of the chapter on page 41 to see how to find a repeated footnote.]

Our interest is in the function $\omega \mapsto x_T(\omega) = x_{T(\omega)}(\omega)$, the location of the probe at the time T. Suppose we consider a landing successful if our probe lands within F feet of the ideal landing site s at the time T it does land. We are then most interested in the probability

$$p_F \stackrel{\text{def}}{=} \mathbb{P}\big(\{\omega \in \Omega : \big\|x_T(\omega) - s\big\| < F\}\big)$$

of a successful landing – its value should influence strongly our decision to launch. Now x_T is just a function on Ω, albeit defined in a circuitous way. We should be able to compute the set $\{\omega \in \Omega : \|x_T(\omega) - s\| < F\}$, and if we have enough information about $\mathbb{P}$, we should be able to compute its probability p_F and to make a decision. This is all classical ordinary differential equations (ODE), complicated by the presence of a parameter ω: straightforward in principle, if possibly hard in execution.

The Obstacle

As long as the *paths* $Z_.(\omega) : s \mapsto Z_s(\omega)$ of the background noise are right-continuous and have finite variation, the integrals $\int \cdots_s dZ_s$ appearing in equations (1.1.6) and (1.1.7) have a perfectly clear classical meaning as Lebesgue–Stieltjes integrals, and Picard's scheme works as usual, under the assumption that the coupling coefficients a, b are Lipschitz functions (see pages 274–281).

Now, since we do not know the background noise Z precisely, we must make a model about its statistical behavior. And here a formidable obstacle rears its head: the simplest and most plausible statistical assumptions about Z force it to be so irregular that the integrals of (1.1.6) and (1.1.7) cannot be interpreted in terms of the usual integration theory. The moment we stipulate some symmetry that merely expresses the idea that we don't know it all, obstacles arise that cause the paths of Z to have infinite variation and thus prevent the use of the Lebesgue–Stieltjes integral in giving a meaning to expressions like $\int X_s \, dZ_s(\omega)$.

Here are two assumptions on the random *driving term* Z that are eminently plausible:

(a) The expectation of the increment $dZ_t \approx Z_{t+h} - Z_t$ should be zero; otherwise there is a *drift* part to the noise, which should be subsumed in the first driving term $\int \cdot \, ds$ of equation (1.1.6). We may want to assume a bit more, namely, that if everything of interest, including the noise $Z_.(\omega)$, was actually observed up to time t, then the future increment $Z_{t+h} - Z_t$ still averages to zero. Again, if this is not so, then a part of Z can be shifted into a driving term of finite variation so that the remainder satisfies this condition – see theorem 4.3.1 on page 221 and proposition 4.4.1 on page 233. The mathematical formulation of this idea is as follows: let $\mathcal{F}_t$ be the σ-algebra generated by the collection of all observations that can be made before and at time t; $\mathcal{F}_t$ is commonly and with intuitive appeal called the *history* or *past* at time t. In these terms our assumption is that the *conditional expectation*

$$\mathbb{E}\left[Z_{t+h} - Z_t \middle| \mathcal{F}_t\right]$$

of the future differential noise given the past vanishes. This makes Z a *martingale on the filtration* $\mathcal{F}_. = \{\mathcal{F}_t\}_{0\le t<\infty}$ – these notions are discussed in detail in sections 1.3 and 2.5.

(b) We may want to assume further that Z does not change too wildly with time, say, that the paths $s \mapsto Z_s(\omega)$ are continuous. In the example of our space probe this reflects the idea that it will not blow up or be hit by lightning; these would be huge and sudden disturbances that we avoid by careful engineering and by not launching during a thunderstorm.

A background noise Z satisfying (a) and (b) has the property that *almost none of its paths $Z_.(\omega)$ is differentiable at any instant* – see exercise 3.8.13 on page 152. By a well-known theorem of real analysis,[2] the path $s \mapsto Z_s(\omega)$ does not have finite variation on any time-interval; and this irregularity happens for almost every $\omega \in \Omega$!

We are stumped: since $s \mapsto Z_s$ does not have finite variation, the integrals $\int \cdots dZ_s$ appearing in equations (1.1.6) and (1.1.7) do not make sense in any way we know, and then neither do the equations themselves.

Historically, the situation stalled at this juncture for quite a while. Wiener made an attempt to define the integrals in question in the sense of distribution theory, but the resulting *Wiener integral* is unsuitable for the iteration scheme (1.1.7), for lack of decent limit theorems.

Itô's Way Out of the Quandary

The problem is evidently to give a meaning to the integrals appearing in (1.1.6) and (1.1.7). Not only that, any prospective integral must have rather good properties: to show that the iterates X^n of (1.1.7) form a Cauchy sequence and thus converge there must be estimates available; to show that their limit is the solution of (1.1.6) there must be a limit theorem that permits the interchange of limit and integral, to wit,

$$\int_0^t \lim_n b(X_s^n)\, dZ_s = \lim_n \int_0^t b(X_s^n)\, dZ_s \,.$$

In other words, what is needed is an integral satisfying the Dominated Convergence Theorem, say. Convinced that an integral with this property cannot be defined *pathwise*, i.e., ω for ω, the Japanese mathematician Itô decided to try for an integral in the sense of the L^2-mean. His idea was this: while the sums

$$S_{\mathcal{P}}(\omega) \overset{\text{def}}{=} \sum_{k=1}^{K} b\big(X_{\sigma_k}(\omega)\big)\left(Z_{s_{k+1}}(\omega) - Z_{s_k}(\omega)\right)\,, \quad s_k \le \sigma_k \le s_{k+1}\,, \tag{1.1.8}$$

[2] See for example [96, pages 94–100] or [9, page 157 ff.].

which appear in the usual definition of the integral, do not converge for any $\omega \in \Omega$, there may obtain *convergence in mean* as the partition $\mathcal{P} = \{s_0 < s_1 < \ldots < s_{K+1}\}$ is refined. In other words, there may be a random variable I such that

$$\| S_{\mathcal{P}} - I \|_{L^2} \to 0 \quad \text{as} \quad \text{mesh}[\mathcal{P}] \to 0 \,.$$

And if $S_{\mathcal{P}}$ should not converge in L^2-mean, it may converge in L^p-mean for some other $p \in (0, \infty)$, or at least in measure ($p = 0$).

In fact, this approach succeeds, but not without another observation that Itô made: for the purpose of Picard's scheme it is not necessary to integrate all processes.[3] *An integral defined for non-anticipating integrands suffices.* In order to describe this notion with a modicum of precision, we must refer again to the σ-algebras $\mathcal{F}_t$ comprising the history known at time t. The integrals $\int_0^t a(X_0)\,ds = a(X_0)\cdot t$ and $\int_0^t b(X_0)\,dZ_s(\omega) = b(X_0)\cdot\big(Z_t(\omega) - Z_0(\omega)\big)$ are at any time measurable on $\mathcal{F}_t$ because Z_t is; then so is the first Picard iterate X_t^1. Suppose it is true that the iterate X^n of Picard's scheme is at all times t measurable on $\mathcal{F}_t$; then so are $a(X_t^n)$ and $b(X_t^n)$. Their integrals, being limits of sums as in (1.1.8), will again be measurable on $\mathcal{F}_t$ at all instants t; then so will be the next Picard iterate X_t^{n+1} and with it $a(X_t^{n+1})$ and $b(X_t^{n+1})$, and so on. In other words, the integrands that have to be dealt with *do not anticipate the future*; rather, they are at any instant t measurable on the past $\mathcal{F}_t$. If this is to hold for the approximation of (1.1.8) as well, we are forced to choose for the point σ_i at which $b(X)$ is evaluated the left endpoint s_{i-1}. We shall see in theorem 2.5.24 that the choice $\sigma_i = s_{i-1}$ permits martingale[4] drivers Z – recall that it is the martingales that are causing the problems.

Since our object is to obtain statistical information, evaluating integrals and solving stochastic differential equations in the sense of a mean would pose no philosophical obstacle. It is, however, now not quite clear what it is that equation (1.1.5) models, if the integral is understood in the sense of the mean. Namely, what is the mechanism by which the random variable dZ_t affects the change dX_t in mean but not through its actual realization $dZ_t(\omega)$? Do the possible but not actually realized courses-of-history[1] somehow influence the behavior of our system? We shall return to this question in remarks 3.7.27 on page 141 and give a rather satisfactory answer in section 5.4 on page 310.

Summary: The Task Ahead

It is now clear what has to be done. First, the stochastic integral in the L^p-mean sense for non-anticipating integrands has to be developed. This

[3] A process is simply a function $Y : (s,\omega) \mapsto Y_s(\omega)$ on $\mathbb{R}_+ \times \Omega$. Think of $Y_s(\omega) = b(X_s(\omega))$.
[4] See page 5 and section 2.5, where this notion is discussed in detail.

is surprisingly easy. As in the case of integrals on the line, the integral is defined first in a non-controversial way on a collection $\mathcal{E}$ of *elementary integrands*. These are the analogs of the familiar step functions. Then that *elementary integral* is extended to a large class of processes in such a way that it features the Dominated Convergence Theorem. This is not possible for arbitrary driving terms Z, just as not every function z on the line is the distribution function of a σ-additive measure – to earn that distinction z must be right-continuous and have finite variation. The stochastic driving terms Z for which an extension with the desired properties has a chance to exist are identified by conditions completely analogous to these two and are called *integrators*.

For the extension proper we employ Daniell's method. The arguments are so similar to the usual ones that it would suffice to state the theorems, were it not for the deplorable fact that Daniell's procedure is generally not too well known, is even being resisted. Its efficacy is unsurpassed, in particular in the stochastic case.

Then it has to be shown that the integral found can, in fact, be used to solve the stochastic differential equation (1.1.5). Again, the arguments are straightforward adaptations of the classical ones outlined in the beginning of section 5.1, jazzed up a bit in the manner well known from the theory of ordinary differential equations in Banach spaces (e.g., [22, page 279 ff.] – the reader need not be familiar with it, as the details are developed in chapter 5). A pleasant surprise waits in the wings. Although the integrals appearing in (1.1.6) cannot be understood pathwise in the ordinary sense, there is an algorithm that solves (1.1.6) pathwise, i.e., ω–by–ω. This answers satisfactorily the question raised above concerning the meaning of solving a stochastic differential equation "in mean."

Indeed, why not let the cat out of the bag: the algorithm is simply the method of Euler–Peano. Recall how this works in the case of the deterministic differential equation $dX_t = a(X_t)\,dt$. One gives oneself a threshold δ and defines inductively an approximate solution $X_t^{(\delta)}$ at the points $t_k \stackrel{\text{def}}{=} k\delta$, $k \in \mathbb{N}$, as follows: if $X_{t_k}^{(\delta)}$ is constructed, wait until the driving term t has changed by δ, and let $t_{k+1} \stackrel{\text{def}}{=} t_k + \delta$ and

$$X_{t_{k+1}}^{(\delta)} = X_{t_k}^{(\delta)} + a(X_{t_k}^{(\delta)}) \times (t_{k+1} - t_k) ;$$

between t_k and t_{k+1} define $X_t^{(\delta)}$ by linearity. The compactness criterion A.2.38 of Ascoli–Arzelà allows the conclusion that the polygonal paths $X^{(\delta)}$ have a limit point as $\delta \to 0$, which is a solution. This scheme actually expresses more intuitively the meaning of the equation $dX_t = a(X_t)\,dt$ than does Picard's. If one can show that it converges, one should be satisfied that the limit is for all intents and purposes a solution of the differential equation.

In fact, the adaptive version of this scheme, where one waits until the *effect of the driving term* $a(X_{t_k}^{(\delta)}) \times (t - t_k)$ is sufficiently large to define t_{k+1}

and $X^{(\delta)}_{t_{k+1}}$, does converge for almost all $\omega \in \Omega$ in the stochastic case, when the deterministic driving term $t \mapsto t$ is replaced by the stochastic driver $t \mapsto Z_t(\omega)$ (see section 5.4).

So now the reader might well ask why we should go through all the labor of stochastic integration: integrals do not even appear in this scheme! And the question of what it means to solve a stochastic differential equation "in mean" does not arise. The answer is that there seems to be no way to prove the almost sure convergence of the Euler–Peano scheme directly, due to the absence of compactness. One has to show[5] that the Picard scheme works before the Euler–Peano scheme can be proved to converge.

So here is a new perspective: what we mean by a solution of equation (1.1.4),

$$dX_t(\omega) = a(X_t(\omega))\,dt + b(X_t(\omega))\,dZ_t(\omega)\ ,$$

is a limit to the Euler–Peano scheme. Much of the labor in these notes is expended just to establish via stochastic integration and Picard's method that this scheme does, in fact, converge almost surely.

Two further points. First, even if the model for the background noise Z is simple, say, is a Wiener process, the stochastic integration theory must be developed for integrators more general than that. The reason is that the solution of a stochastic differential equation is itself an integrator, and in this capacity it can best be analyzed. Moreover, in mathematical finance and in filtering and control theory, the solution of one stochastic differential equation is often used to drive another.

Next, in most applications the state of the system will have many components and there will be several background noises; the stochastic differential equation (1.1.5) then becomes[6]

$$X^\nu_t = C^\nu_t + \sum_{1\le\eta\le d} \int_0^t F^\nu_\eta[X^1,\ldots,X^n]\,dZ^\eta\ , \qquad \nu = 1,\ldots,n\ .$$

The state of the system is a vector $X = (X^\nu)^{\nu=1\ldots n}$ in $\mathbb{R}^n$ whose evolution is driven by a collection $\{Z^\eta : 1 \le \eta \le d\}$ of scalar integrators. The d vector fields $F_\eta = (F^\nu_\eta)^{\nu=1\ldots n}$ are the *coupling coefficients*, which describe the effect of the background noises Z^η on the change of X. $C_t = (C^\nu_t)^{\nu=1\ldots n}$ is the *initial condition* – it is convenient to abandon the idea that it be constant. It eases the reading to rewrite the previous equation in vector notation as[7]

$$X_t = C_t + \int_0^t F_\eta[X]\,dZ^\eta\ . \tag{1.1.9}$$

[5] So far – here is a challenge for the reader!
[6] See equation (5.2.1) on page 282 for a more precise discussion.
[7] We shall use the ***Einstein convention*** throughout: summation over repeated indices *in opposite positions* (the η in (1.1.9)) is implied.

The form (1.1.9) offers an intuitive way of reading the stochastic differential equation: the noise Z^η *drives* the state X in the direction $F_\eta[X]$. In our example we had four driving terms: $Z^1_t = t$ is time and F_1 is the systemic force; Z^2 describes the gravitational gullies and F_2 their effect; and Z^3 and Z^4 describe the gusts of wind and the gaggle of geese, respectively. The need for several noises will occasionally call for estimates involving whole slews $\{Z^1, ..., Z^d\}$ of integrators.

1.2 Wiener Process

Wiener process[8] is the model most frequently used for a background noise. It can perhaps best be motivated by looking at Brownian motion, for which it was an early model. Brownian motion is an example not far removed from our space probe, in that it concerns the motion of a particle moving under the influence of noise. It is simple enough to allow a good stab at the background noise.

Example 1.2.1 (Brownian Motion) Soon after the invention of the microscope in the 17th century it was observed that pollen immersed in a fluid of its own specific weight does not stay calmly suspended but rather moves about in a highly irregular fashion, and never stops. The English physicist Brown studied this phenomenon extensively in the early part of the 19th century and found some systematic behavior: the motion is the more pronounced the smaller the pollen and the higher the temperature; the pollen does not aim for any goal – rather, during any time-interval its path appears much the same as it does during any other interval of like duration, and it also looks the same if the direction of time is reversed. There was speculation that the pollen, being live matter, is propelling itself through the fluid. This, however, runs into the objection that it must have infinite energy to do so (jars of fluid with pollen in it were stored for up to 20 years in dark, cool places, after which the pollen was observed to jitter about with undiminished enthusiasm); worse, ground-up granite instead of pollen showed the same behavior.

In 1905 Einstein wrote three Nobel-prize–worthy papers. One offered the Special Theory of Relativity, another explained the Photoeffect (for this he got the Nobel prize), and the third gave an explanation of Brownian motion. It rests on the idea that the pollen is kicked by the much smaller fluid molecules, which are in constant thermal motion. The idea is not, as one might think at first, that the little jittery movements one observes are due to kicks received from particularly energetic molecules; estimates of the distribution of the kinetic energy of the fluid molecules rule this out. Rather, it is this: the pollen suffers an enormous number of collisions with the molecules of the surrounding fluid, each trying to propel it in a different direction, but mostly canceling each other; *the motion observed is due to*

[8] "Wiener process" is sometimes used without an article, in the way "Hilbert space" is.

statistical fluctuations. Formulating this in mathematical terms leads to a stochastic differential equation[9]

$$\begin{pmatrix} dx_t \\ dp_t \end{pmatrix} = \begin{pmatrix} p_t/m\,dt \\ -\alpha\,p_t\,dt\,+\,d\boldsymbol{W}_t \end{pmatrix} \tag{1.2.1}$$

for the location (x,p) of the pollen *in its phase space.* The first line expresses merely the definition of the momentum p; namely, the rate of change of the location x in $\mathbb{R}^3$ is the velocity $v = p/m$, m being the mass of the pollen. The second line attributes the change of p during dt to two causes: $-\alpha p\,dt$ describes the resistance to motion due to the viscosity of the fluid, and $d\boldsymbol{W}_t$ is the sum of the very small momenta that the enormous number of collisions impart to the pollen during dt. The random driving term is denoted $\boldsymbol{W}$ here rather than Z as in section 1.1, since the model for it will be a Wiener process.

This explanation leads to a plausible model for the background noise $\boldsymbol{W}$: $d\boldsymbol{W}_t = \boldsymbol{W}_{t+dt} - \boldsymbol{W}_t$ is the sum of a huge number of exceedingly small momenta, so by the Central Limit Theorem A.4.4 we expect $d\boldsymbol{W}_t$ to have a normal law. (For the notion of a law or distribution see section A.3 on page 391. We won't discuss here Lindeberg's or other conditions that would make this argument more rigorous; let us just assume that whatever condition on the distribution of the momenta of the molecules needed for the CLT is satisfied. We are, after all, doing heuristics here.)

We do not see any reason why kicks in one direction should, on the average, be more likely than in any other, so this normal law should have expectation zero and a multiple of the identity for its covariance matrix. In other words, it is plausible to stipulate that $d\boldsymbol{W}$ be a 3-vector of identically distributed independent normal random variables. It suffices to analyze one of its three scalar components; let us denote it by dW.

Next, there is no reason to believe that the total momenta imparted during non-overlapping time-intervals should have anything to do with one another. In terms of W this means that for consecutive instants $0 = t_0 < t_1 < t_2 < \ldots < t_K$ the corresponding family of consecutive ***increments***

$$\left\{ W_{t_1} - W_{t_0}\,,\, W_{t_2} - W_{t_1}\,,\, \ldots\,,\, W_{t_K} - W_{t_{K-1}} \right\} ,$$

should be independent. In self-explanatory terminology: we stipulate that W have ***independent increments.***

The background noise that we visualize does not change its character with time (except when the temperature changes). Therefore the law of $W_t - W_s$ should not depend on the times s,t individually but only on their difference, the elapsed time $t - s$. In self-explanatory terminology: we stipulate that W be ***stationary.***

[9] Edward Nelson's book, *Dynamical Theories of Brownian Motion* [82], offers a most enjoyable and thorough treatment and opens vistas to higher things.